計畫編號: NSC 89-2118-M-032-019
執行期限: 89 **年** 8 **月** 1 **日至** 90 **年** 7 **月** 31 **日**
主持人: **鄭惟厚　　淡江大學數學系**
計畫名稱: **隨機性的無母數檢定**

中文摘要: **利用鄭惟厚和陳奕良所導出的尖峰及連升的確實聯合分布**, **考慮檢定隨機性的新統計量**, **並檢視其檢力**.
關鍵詞: **隨機性**,**無母數**,**檢定**
Abstract: We consider the problem of nonparametric test for randomness. By using the results of Chen & Cheng, we intend to come up with new test(s).
Keywords: randomness, nonparametric, test

計畫緣由與目的:
隨機性檢定是常用且重要的問題, **是否能提出新且好的無母數檢定是值得探討的題目**. **原已有論文討論過用尖峰**(peak) **或 連升**(rise) **來檢定**, **因為陳奕良和鄭惟厚導出尖峰和連升的確實聯合分布**, **因此想考慮用尖峰**, **連升**, **凹點及連降的函數來找出新的檢定統計量**.

結果與討論
考慮 $\mathrm{I}_n = \{1,2,....,n\}$,**而** S **為所有** $\{1,2,\ldots,n\}$ **可能的排列**(permutation)**所成的集合**.**這些排列具有均勻分布**(uniform distribution),**及每個排列之概率為** $\frac{1}{n!}$.**若** I_n **的一個隨機排列記為** $\Pi_n = (®_1, ®_2, \ldots\ldots ®_n)$,**則** $®_i$ **稱為此隨機排列** Π_n **的元素**(element),**而我們依此元素** $®_i$ **與其左右兩個元素** $®_{i-1}$ **與** $®_{i+1}$ **之大小關係**,**定義以下四種情況** :

(i) **若元素** $®_i$ **滿足** $®_{i-1}, < ®_i > ®_{i+1}$,**則稱為尖峰**(peak),
(ii) **若元素** $®_i$ **滿足** $®_{i-1}, < ®_i < ®_{i+1}$,**則稱為連升**(double rise),
(iii) **若元素** $®_i$ **滿足** $®_{i-1}, > ®_i < ®_{i+1}$,**則稱為凹點**(trough),
(iv) **若元素** $®_i$ **滿足** $®_{i-1}, > ®_i > ®_{i+1}$,**則稱為連降**(double fall).

在討論環狀排列(circular permutation)**的時候**,**首尾元素相連**,**一般技巧上皆會將** $®_{n+1}$ **視為** $®_1$ (**即** $®_{n+1} = ®_1$),**而** $®_0$ **視為** $®_n$ (**即** $®_0 = ®_n$),**如此一來**,**在一個排列裏的每一個元素都可以被歸類成上面所述的**(i) ~(iv)**四種狀況其中的一種**.**我們定義在一個排列** Π_n **裏** ,**其尖峰**,**連升**, **凹點**,**及連降等這四種情況所發生的個數**,**分別用以下符號表示** :

$$\mathrm{A}_n = \sum_{i=1}^{n} \mathrm{I}\{®_{i-1} < ®_i > ®_{i+1}\},$$

$$B_n = \sum_{i=1}^{n} \mathrm{I}\{\text{®}_{i-1} < \text{®}_i < \text{®}_{i+1}\},$$

$$C_n = \sum_{i=1}^{n} \mathrm{I}\{\text{®}_{i-1} > \text{®}_i < \text{®}_{i+1}\},$$

$$D_n = \sum_{i=1}^{n} \mathrm{I}\{\text{®}_{i-1} > \text{®}_i > \text{®}_{i+1}\}.$$

式中的 I{.}**為指標函數** (indicator function).
在環狀情況下,**由於尖峰與凹點是交替出現**,**而且也因為元素**®_1 **與**®_n **首尾相連**,
所以,$A_n = C_n$,**又因為**,$A_n + B_n + C_n + D_n = n$, $n \geq 2$.
所以我們討論 A_n,B_n,C_n,D_n **之聯合分布**(joint distribution),**只需考慮尖峰** A_n **與連升** B_n **兩個變數即可**.

Entringer[1969]**研究尖峰** A_n **上的一些重要結果**.
他用偏微分方程(partial differential equation)**的方法來解出尖峰** A_n **的衍生函數**,
然後再將它展開,**比較係數**,**最後得到尖峰** A_n **的確實分布**.
陳奕良與鄭惟厚憑藉著 Entringer **的結果**,**推導出**(A_n, B_n)**的確實聯合分布**.
有關於尖峰 A_n **的論文已有不少**,**其中包括用尖峰來作隨機性**(randomness)**的檢定**.
由於(A_n, B_n)**的確實聯合分布已被導出**,**因此我們考慮用其他** A_n, B_n **之函數來作隨機性檢定**.

在考慮過一些統計量並用電腦做 Monte Carlo **模擬比較之後**,**較為可行的似乎是** R_n=rise **的個數**-fall **的個數**= $B_n - D_n$.**要找出** R_n **在原始假設** H_0(**即** $X_1, X_2, \ldots X_n$ **為隨機排列**)**下之分布**,**必須用到陳奕良與鄭惟厚之結果**,**現敘述如下**:

令 $\Psi(n,a,b) = n! \cdot P_r[A_n = a, B_n = b]$, $\Psi(n,a,.) = n! \cdot P_r[A_n = a]$, **則有**
對於 $n \geq 2$,

$$\Psi(n;a,b) = \binom{n-2a}{b}\left(\frac{1}{2}\right)^{n-2a} \Psi(n,a,.),$$

式中的 a=1,2…..,$[\frac{n}{2}]$,b=0,1,….,n-2a
而此處的$\Psi(n,a,.)$ **是根據** Entringer **的結果**, **即當** $n \geq 2$,$a \geq 0$ **時**.

$$\Psi(n,a,.) = (-1)^{n+a} \cdot n \cdot 2^{n-1} \cdot \sum_{r=a-1}^{[\frac{n}{2}]} \sum_{t=n-2r-1}^{n-1} \frac{(-1)^t \cdot t!}{2^t} \binom{r}{a-1}\binom{t-1}{n-2r-2} s(n-1,t),$$

此式中的 s(n-1,t)=$\frac{(-1)^t}{t!}\sum_{s=1}^{t}(-1)^s \binom{t}{s} s^{n-1}$ **為第二類史特林數**(Stirling number of the second kind).

根據上面結果,可得 R_n **在** H_0 **之下之確實分布,例如,當** n=10 **時,我們得到如下表所列的結果,**

R_n **的值**	**機率**
-8	0.000003
-6	0.0014
-4	0.0402
-2	0.2430
0	0.4304
2	0.2431
4	0.0403
6	0.0014
8	0.000003

表列之值有四捨五入之誤差,否則應為對 0 **對稱之分布. 因為(** A_n, B_n**)的確實聯合分布的式子非常複雜, 所以樣本數較大時, 要計算任何** A_n, B_n **之函數的分布時, 其過程是相當繁瑣的.**

模擬檢定隨機對應上升趨勢(trend) **的情況, 分二種情形:**

(i) H_0 : $X_1, X_2, \ldots X_{20}$ i.i.d N(0,2) v.s H_1 : $X_i \sim N(\frac{i}{5}, 2)$,i=1,2,….20

(ii) H_0 : $X_1, X_2, \ldots X_{40}$ i.i.d N(0,1) v.s H_1 : $X_i \sim N(\frac{i}{5}, 1)$,i=1,2,….40

各做 2000 **組,分別得到以下的結果(列在下一頁).**

以 (ii) **為例**, n=40 **時**, **以** $R_n \geq 8$ **為棄卻域時**, **~約為** .03, **而在** trend alternative **之下**, power **只有** .1547, **不理想. 若以** $R_n \geq 6$ **為棄卻域**, **~約為** .0755, **而檢力為** .3182.

成果自評

試過好些個統計量都沒有看到我想見到的結果,模擬過程頗為費時, 因此還有一些模擬還來不及做, 只好繼續努力吧.

參考文獻 (**只列最直接且重要者**)

Entringer, R.C.(1969). Enumeration of permutations of (1,…,n) by number of maxima. Duke Math. J. 36,575-579.

陳奕良. (1999). **關於數列波動的一些機率結果. 博士論文**

H_1 之下的模擬結果

Case (i)

R_n 的值	出現的次數	機率
-12	0	0.0000
-10	0	0.0000
-8	1	0.0005
-6	26	0.0130
-4	147	0.0735
-2	362	0.1810
0	601	0.3005
2	540	0.2700
4	233	0.1165
6	82	0.0410
8	7	0.0035
10	1	0.0005
12	0	0.0000

Case (ii)

R_n 的值	出現的次數	機率
-18	0	0.0000
-16	0	0.0000
-14	0	0.0000
-12	0	0.0000
-10	1	0.0005
-8	2	0.0010
-6	17	0.0085
-4	53	0.0265
-2	140	0.0700
0	330	0.1650
2	436	0.2180
4	381	0.1905
6	327	0.1635
8	187	0.0935
10	83	0.0415
12	34	0.0170
14	5	0.0025
16	2	0.0001
18	2	0.0001